

detector indicating that the distance between the locations detected for each object is less than the threshold distance.

10. A system according to any preceding clause, wherein the object detector comprises a cascade classifier trained with images of passive non-luminous objects.

11. A system according to any preceding clause, wherein the object pose detector comprises a machine learning model trained to perform six-dimensional pose estimation for each passive non-luminous object detected being held by the user in the obtained images.

12. A system according to clause 11, wherein the obtained images comprise at least two passive non-luminous objects being held by a user;

[0120] wherein the object pose detector is configured to detect a plurality of keypoints for each passive non-luminous object being held by the user; and

[0121] wherein the user input generator is configured to generate a user input based on a detection of the distance between at least some of the keypoints for each object being detected as greater or less than a threshold distance.

13. A system according to clause 11 or clause 12, wherein the obtained images comprise at least two passive non-luminous objects being held by different respective users;

[0122] the system further comprising a user identification unit operable to associate each passive non-luminous object with a different respective user based on at least one of a relative distance and depth between the passive non-luminous objects exceeding a threshold value; and

[0123] wherein the user input generator is configured to generate user inputs for each user participating in a video game session based on the changes in pose detected of the corresponding passive non-luminous objects.

14. A system according to any of clauses 11 to 13, wherein the user input generator is configured to generate a different respective user input for each change in pose in a respective dimension.

15. A system according to any preceding clause, further comprising:

[0124] an image generator operable to generate an image of a virtual button for superimposing on top of a user's view of the passive non-luminous object;

[0125] a display for displaying the image of the virtual button at a location that corresponds to a location on the surface of the passive non-luminous object;

[0126] a finger detector operable to detect a user's finger in the obtained images and a location of the finger relative to the passive non-luminous object;

[0127] wherein the finger detector is configured detect when a user's finger coincides with the location of the virtual button, and in response to determining that the user's finger coincides with the location of the virtual button, provide an input to the user input generator; and

[0128] wherein the user input generator is configured to generate a user input in accordance with a pressing of the virtual button.

16. A system according to any preceding clause comprising a camera operable to capture images of the non-luminous object being held by the user; and

[0129] wherein the input unit is operable to receive the images of object captured by the camera.

17. A system according to any preceding clause, comprising the video game unit operable to receive the input from the user input generator.

18. A method for generating user inputs for a video game, the method comprising:

[0130] obtaining a plurality of images of a passive non-luminous object being held by a user;

[0131] detecting pixels in the obtained images as corresponding to the passive non-luminous object, the pixels corresponding to the object itself and not a physical identifier that has been added to the object;

[0132] detecting changes in pose of the passive non-luminous object based on the obtained images of the object;

[0133] wherein detecting the pose of the passive non-luminous object is based on at least one of a (i) contour detection operation and (ii) the output of a machine learning model that has been trained to detect the poses of objects in images;

[0134] generating, based on the detected changes in pose of the passive non-luminous object, a user input for controlling a virtual object in a video game; and

[0135] transmitting the generated user input to a video game processor so as to control the virtual object in a video game in accordance with the generated user input.

19. A method according to clause 18, wherein detecting pixels in the obtained images as corresponding to the passive non-luminous object comprises detecting a contour in obtained images as corresponding to the passive non-luminous object being held by the user; and

[0136] wherein detected changes in pose of the passive non-luminous object comprises detecting changes in at least one of the orientation, position and area of the contour in the obtained images.

20. A method according to clause 19, wherein detecting pixels in the obtained images as corresponding to the passive non-luminous object comprises detecting a plurality of contours in the obtained images and identifying the largest contour as corresponding to the passive non-luminous object being held by the user.

21. A method according to clause 19 or clause 20, comprising filtering one or more pre-determined colours from the obtained images prior to detecting a contour for the passive non-luminous object being held by the user.

22. A method according to any of clauses 19 to 21, comprising detecting at least two passive non-luminous objects being held by the user, each passive non-luminous object being detected based on one or more contours detected as corresponding to that object;

[0137] identifying a location within the image representation of each object based on the one or more contours detected for that object; and

[0138] detecting a change in pose of the at least two passive non-luminous object based on changes in orientation of a line intersecting the detected locations, and generating a directional user input based on changes in orientation of the line intersecting the locations relative to a default orientation of said line with respect to the detected locations.

23. A method according to any of clauses 18 to 22, wherein estimating the pose comprises inputting the obtained images to a machine learning model trained to perform six-dimensional pose estimation.

24. A method according to any of clauses 18 to 23, comprising detecting at least two passive non-luminous objects in the obtained images, the method further comprising:

[0139] detecting at least two users in the obtained images;

[0140] determining a relative distance between each user and each object;